

The Dialects of Old Indo-Aryan

M. B. Emeneau

UNIVERSITY OF CALIFORNIA, BERKELEY

The most obvious fact about Sanskrit, or Old Indo-Aryan (OIA), when we are interested in its dialects, is that the records show large corpora in what are essentially two rather divergent dialects—Vedic and classical. But, as soon as this statement is made, it is seen to be too simple. Classical Sanskrit is a literary language written according to the book—that is, Pāṇini's grammar, and following it more or less correctly. We find in it no dialects, no chronological development, except loss and at times invasion from the vernaculars of the users, and no geographical divergences. Vedic Sanskrit, however, is different. It is anything but a unified language, a language of one dialect only. It shows even within the oldest member of the corpus, the Ṛgveda, linguistic features that can be explained only by positing their origin in slightly differing dialects, and within the total Vedic corpus there is a sliding scale of clusters of dialectal features that run all the way from those that are most different from classical Sanskrit to those that are, in fact, taken by most scholars in the field to be essentially the dialect that Pāṇini described as his norm. No dissent from this sketch need be caused by Pāṇini's label of "Vedic" for many forms that differ from those of his norm. It is the prose of the latest Brāhmaṇa texts and the Upaniṣads that is nearest to his language. His "Vedic" variants are those of the earlier, essentially "canonical" collections.¹

Pāṇini described, presumably, a dialect of Northwestern India (his birthplace Śālātura was there), and he is dated in the middle of the fifth century B.C., with the usual plus-or-minus of Indian chronology, this time perhaps plus or minus a hundred and fifty years.² Besides his references to Vedic forms, he mentions geographical variants, but not so that we can do much in the way of identifying dialects. All that we are really told is that there were other dialects in North India besides the one he described. As has been said by L. Renou,³ his information allows us to

¹ So I understand P. Thieme to hold in *Pāṇini and the Veda* (Allahabad, 1935).

² V. S. Agrawala, *India as Known to Pāṇini* 475 (Lucknow, 1953), for the attempt at precision; M. B. Emeneau, *JAOS* 75.146 n. 3 (1955) for a very short summary of views.

³ *Histoire de la langue sanskrite* 67 (Lyon, 1956).

draw no isoglosses. And the same is true of what little dialectal information we are given by the two great commentators Kātyāyana and Patañjali (the latter in the second century B.C.).

A very crucial point in this matter of dialects is the relationship between classical Sanskrit and the extreme Vedic dialect of the Ṛgveda. A chronological difference, one of from five hundred years up, is the first matter to be mentioned. In saying five hundred years it is possible that there is exaggeration, but I incline to think that Burrow's, Renou's, and others' guess of circa 1200–1000 B.C. for the period of the composition of the hymns of the Ṛgveda is not far out. In a period of five hundred years we could expect much linguistic change to have taken place, if Ṛgvedic Sanskrit were the genetic origin of classical Sanskrit. It is worthwhile to assume for a moment that that is the relation between the two and to identify some of the changes.

Some of the differences that we might superficially hit upon turn out to be something else. For example, as we usually learn Sanskrit, whether it is Westerners or students in India, classical Sanskrit has a strangely indeterminate sort of stress accent and Vedic Sanskrit is characterized by a pitch-accent system. However, Pāṇini's Sanskrit had the same pitch accent as Vedic Sanskrit, and the learning of classical Sanskrit without the pitch accent, which is really what the uncertain stress accent amounts to, results from an invasion of classical Sanskrit by the Middle Indo-Aryan (MIA) and Modern Indo-Aryan (NIA) stress-accent systems which are a historical development diverging from the OIA state of affairs. Similarly, the general equation in meaning of the three past tenses, imperfect, perfect, and aorist, in classical and epic Sanskrit is certainly different from the Ṛgvedic state of affairs (even if we are not completely certain of the Vedic distinctions between these three tenses in all instances). But it is fairly certain that Pāṇini's Sanskrit had the same meanings for these tenses (or perhaps we should prefer to call them tense aspects) that the Veda had. The difference that we see in the epic and classical texts should be characterized as loss during the long period of the use of Sanskrit as a literary language. I would assume that it came about because of the invasion of the literary language by vernacular speech habits of the kind seen in our oldest voluminous MIA material, that in Pali. In this language there had evolved from the OIA system (the Vedic and that described by Pāṇini) a verbal system with one past tense only, formally an amalgam of imperfect and aorist, with very few forms descended from the perfect. In the Prakrits there is much less even than this, and passive constructions using the past participle are usual (when the present is not the narrative tense), prefiguring what is found in the modern vernaculars. It is from the oldest MIA stages, then, that the use of the narrative tenses in epic and classical Sanskrit literature derives.

Some changes may, however, be identified that might be quite straight-line developments from Vedic to classical. Most involve simplification or loss of one sort or another. A straightforward example is loss of the subjunctive except for the retention of first person forms as part of the imperative paradigm. Another is loss of augmentless aorist and imperfect forms in modal use except for the occurrence of such aorist forms with *mā* in prohibitions; they are completely lost in indicative use, with a very few epic exceptions, which are often treated as lapses from the normative grammar, but which in fact need a different treatment. Simplification in syntax is seen in the classical language's treatment of prepositions and verb forms as inseparable compounds, whereas in the Vedic language they had been freely separable.

Phonologically, some such simplifications are identifiable, in spite of the unhappy historical accident that the Ṛgveda, and all other archaic Vedic texts as well, were written down only late within the Vedic period by speakers of what was apparently, for all that is relevant here, classical Sanskrit, and as if the old Vedic texts were in many details, especially of phonology, classical Sanskrit. Reconstruction of the Vedic text with the aid of the meter has allowed the recovery of the law that we know as Sievers-Edgerton's,⁴ with all that this involves of an allomorphy much more complex than that of classical Sanskrit.

In fact, as has been amply demonstrated, classical Sanskrit contains so little of this allomorphy that on the basis of it alone the Vedic state of affairs could not be more than suspected. It is certain that the system of the classical language had already come into existence before the Ṛgvedic corpus was composed in the condition in which we have it, and that this was prior to the writing-down. This condition involves two forms of the text, one by complete utterances (*saṃhitā-pāṭha*) and one with the utterances analyzed into separate words (*pada-pāṭha*). A considerable body of linguistic analysis is presupposed by this edition of the corpus. Such analysis, which states in all detail the relationship between the two forms of text, is provided by the Ṛgveda-prātiśākhya treatise. Its date is, as usual, unknown, and even its relative chronology within the sequence of grammatical works as a whole is not certain. However, informed scholars, with such a notable exception as Thieme, place it prior to Pāṇini, but later, though how much later is very uncertain, than the two text forms of which it states the relationship. As so often in Indian literary scholarship, it is thought that the present treatise is the end result of a development from less developed to more developed and that the original treatise, which is presupposed by the Ṛgveda text forms, was much enlarged to reach the present form and has itself been lost.

⁴ F. Edgerton, "Sievers's Law and IE. Weak-grade Vocalism," *Lg.* 10.235-265 (1934); "The Indo-European Semivowels," *ibid.* 19.83-124 (1943).

The point of going into the matter of the *prātisākhya* analysis is (1) that this analysis is of a Ṛgvedic text which was, at the time of the first exegesis, that is, that of the word-by-word text form, already recited, whether in ritual use or otherwise, with classical Sanskrit phonology insofar as Sievers-Edgerton's law is concerned, but (2) that the *prātisākhya* analysts had some sort of knowledge of the allophony of semivowels that is stated in the law, even if recitation, both then and in all later times in India, ignored the matter.⁵ How was this knowledge derived? Was it something handed down by tradition, or was it theory based on the Vedic practitioners' study of metrics, as Edgerton suggests? That some fragments of tradition remained is at least suggested by the collection of written forms with *iy* and *uw* allophones from Vedic texts later than the Ṛgveda.⁶ Whichever it was, we must of course regret that more was not said or recorded by the *prātisākhya* scholars, when we consider how much of the detail involved in the law is still uncertain. For example, the form written *pūrva-* is never pronounced **pūrwa-* according to the law, for the reason that it is descended from **pṛva-*, or whatever else is required or suggested by the use of laryngeals. Is it possible that it was actually pronounced **pṛva-* (or what not) in the period of the composition of the Ṛgvedic hymns? Certainly, while Sievers-Edgerton's law was in full working order, something other than *pūrva-* must have been the phonemic form; this means a time very close to the period of our hymns. To sum up, this phonological simplification is going from the Ṛgveda to classical Sanskrit, a simplification that is generally the result of analogical spread for each morpheme of one allomorph at the expense of the other, would provide a chronological isogloss, relative rather than absolutely datable, if our assumption were tenable that classical Sanskrit is genetically derivable from Ṛgvedic Sanskrit.

This assumption, however, is not tenable. The contrary evidence is familiar, and the hypothesis drawn therefrom. The evidence includes such phenomena as the Ṛgvedic correspondence of retroflex *ḷ* where classical Sanskrit has *ḍ* intervocalically. Ṛgvedic has innovations in the noun declension, such as the nom. pl. masc. in *-āsas*, which contrasts with *-ās*, which latter is found also in the Ṛgveda and is the only form in the classical language. It has been demonstrated that, in the Ṛgveda, *-āsas* is the form proper to its dialect, from metrical considerations, viz. that written *-ās* many times has to be read *-āsas* to mend defective meter,

⁵ For the *prātisākhya* passages (RPr. 8.22, 17.14 [ed. Regnier]), J. Wackernagel, *Altindische Grammatik* 1².202 (with much bibliography), and Edgerton, *loc. cit.*, esp. *Lg.* 19.92 n. 24.

⁶ As given by Wackernagel 1.200 f.

while *-āsas* never has to be read *-ās*.⁷ Since Iranian also shows entirely parallel phenomena in archaic and archaizing Avestan *-āṇhō* and Old Persian *-āha*, with also *-ā* and *-ā* respectively in these languages, it follows that for both branches of Indo-Iranian it must be assumed that there were dialects that retained the monosyllabic ending, which is guaranteed as Indo-European by Oscan *-ūs*, Umbrian *-us*, Gothic *-ōs* (but it presents difficulties as, e.g., Prokosch presents them), Old Irish *-u* (vocative: *firu* 'men!', etc.). Such an assumption of a dialect with at least several features that are more archaic than the corresponding features in the R̥gvedic dialect is necessary to explain for Indo-European *o*-stems the classical instr. pl. *-āis* beside R̥gvedic *-ebhis* and *-āis*. Again, Iranian has both forms, this time Old Persian only *-aibiš* and Avestan only *-āiš*. The R̥gvedic (and Old Persian) form is for these nouns an innovation found nowhere else in Indo-European; classical *-āis* has congeners in Italic, Greek, Lithuanian.

The R̥gvedic dialect, then, is clearly not the direct ancestor of classical Sanskrit. There must have been, even on this much evidence, several closely related dialects in the period of the R̥gveda composition, one of which is the basic dialect of this text, another of which is basically the ancestor of the classical language of some centuries later. But it is also clear that the R̥gvedic linguistic norm, even apart from hymns that represent something very close to the classical language, was a mixed dialect, and that one of the elements in the mixture was something near to classical Sanskrit. Only such a hypothesis will explain such mixtures as those of *-āsas* and *-ās*, of *-ebhis* and *-āis*, of the instr. sg. endings *-ā* and *-ena* of IE *o*-stems, of the neuter plurals of Indo-European *o*-stems in *-ā* and *-āni*, and so on. The composers of the older R̥gveda hymns, that is, not those that are for all practical purposes in classical Sanskrit, must have been speakers of many dialects, and the R̥gvedic language represents a mixture, probably of most of them; one very like, or even identical with, the ancestor of classical Sanskrit probably formed one very important element in the mixture. We have an inkling, therefore, so far as direct evidence goes, of the oldest Indo-Aryan of North India as a large dialect area whose speakers were unified by a common culture and by the religion that provides us with the evidential documents; there probably were other dialects as well, outside of this social and religious milieu, as we shall see. The dialects seem to have been subject to interborrowing, even the one which in theory was most free from such borrowing because it was the sacrosanct vehicle of the holy texts. It was a dialect area subject to typical Schmidtian "wave" effects. No other reconstruction will explain the R̥gvedic dialect as we know it, nor its relations with the

⁷ So Wackernagel 3.100, following C. R. Lanman, "Noun Inflection in the R̥gveda," *JAOS* 10 (1878), esp. 345.

Sanskrit of Pāṇini. If it is impossible to be very certain about much of this, the chronological differences in the records account for some of the uncertainty, and the many gaps in the record, both synchronic and diachronic, account for more.

The earliest documents that we have in MIA are the Aśokan inscriptions of the middle of the third century B.C. This date, though post-Pāṇini, is prior to the whole corpus of the classical Sanskrit literature. It should not be surprising, then, to find words of MIA origin in the classical literature, even if their numbers are not excessive and if some of them are of fairly late attestation or only lexical. Exx. *masṛṇa-* 'soft, smooth, tender, mild' (kāvyā) < Pkt. *masaṇa-*, *masiṇa-* < Skt. *mṛtsna-* or *mārtsna-* 'ground fine or small'; *mārṣa*, *māriṣa* 'friend!' (epic, etc.; *mārṣa* in Bhāsa's dramas, see Emeneau, *Lg.* 39.104 [1963]) < Pkt. *māriṣa-* < Skt. *mādrṣa-* 'like me'; *prāgbhāra-* 'mountain slope' (kāvyā) < Pkt. *pabbhāra-* 'sloping' (cf. also BHS *prāgbhāra-* id.⁸) < Skt. **prahvāra-*, cf. *hṛṣ-* 'to bend, slope', probably *prahva-* 'sloping'; and *prāgbhāra-* 'mass, multitude, heap' (late kāvyā) < Pkt. *pabbhāra-* id. < Skt. *pra + bhāra-*.⁹

But it is strictly relevant to our interest to determine how far back in the OIA record such borrowings from MIA can be attested, since, in default of other records, such borrowings are almost the only evidence available on the relative chronology of the emergence of MIA linguistic traits. To be sure, we have already pointed to several features of classical Sanskrit as being probably MIA in origin. One of these is the merging of the old past tenses into one meaning in epic and classical Sanskrit. Pāṇini still knew the differences. There has been discussion whether the epic in fact does not have an origin that is pre-Pāṇinian in time. If it had, it would follow that its use of the tenses might have derived from a MIA tendency already existing before Pāṇini. But there are too many gaps in the evidence for this to be anything but speculation.

Phonological Middle-Indicisms in the Vedic texts later than the R̥gveda have already been pointed out by many scholars, in greatest mass by Bloomfield and Edgerton, *Vedic Variants* 2.20–25 and *passim*. Some of the material collected there is open to uncertainty,¹⁰ but not all,

⁸ B(uddhist) H(ybrid) S(anskrit) material must, of course, not be treated as Sanskrit, but as MIA.

⁹ Emeneau, *Lg.* 36.541 (1960).

¹⁰ The most striking and puzzling example is perhaps *avatā-* 'a well', three times in R̥gveda passages and replaced by *avaṭa-* in the Sāmaveda and the Taittirīya Yajurveda. The *Vedic Variants* interpretation is followed by Mayrhofer. Wackernagel, 1.167, takes *avaṭa-* to be a Middle-Indicisms from **avar* 'down,' with *rt > ṭ*. Even though Burrow (*The Sanskrit Language* 129 [London, 1955]) would refine this and posit an *r/n* formation (**avṛ-ta-* > *avata-*; **avṛ-ta-* > *avaṭa-*), this ignores the fact that *avaṭa-* is only post-R̥gvedic and may well be secondary to *avata-* and not

and it seems to follow clearly enough that there were dialects that were contemporary with at least later Vedic texts and that had phonological features that we know from the later attested MIA dialects.

It will be more satisfactory, however, to pinpoint the matter as sharply as possible. In the first place, there is some clear evidence in Pāṇini. This we should expect because of his chronological proximity to Aśoka and, even more, to the Buddha. The latter's doctrine was that teaching of religion should take place in the colloquial dialects rather than in "Vedic."¹¹ This, together with the historical outcome, viz. Aśoka's Prakrits, Pali, Gāndhārī Prakrit, Buddhist Hybrid Sanskrit, all used as vehicles of Buddhist doctrine, means that in the Buddha's time there were MIA dialects. One of Pāṇini's words is *māireya-* 'an intoxicating drink' (6.2.70). It is from Pkt. *māireya-*, which is an extension derived from Skt. *madira-* id. (Ṛgveda +). As has been pointed out by earlier scholars, loss of OIA intervocalic *d* (and other voiced stops) is attested in general only in Māhārāṣṭrī Prakrit. Pāṇini's word is attestation of a dialect with this feature centuries earlier than we would otherwise expect it.

It is hardly profitable to go into details about Middle-Indicisms in the later Vedic texts. Identifications of such in the Ṛgveda are of chief interest for the chronology we wish to establish. Some scholars have been rather unwilling to admit that they are to be found in this oldest corpus. Renou¹² seems very reluctant, and writes of "justes remarques de Mansion contre l'hypothèse d'un prākṛit contemporain du RV." In translating the introduction to the 1957 edition of Wackernagel's *Altindische Grammatik* Renou had to write (p. 7) of the existence of a MIA dialect "dès l'époque des Hymnes," "une langue . . . qui portait en elle les traits principaux de la phase la plus ancienne du moyen indien, ce qu'on appelle le niveau pāli." But he is here only translating Wackernagel, and in note 80, which gives a large bibliography pro and con, he again refers to Mansion, "qui proteste avec raison contre l'idée . . . d'un pkt contemporain du RV." The bibliography (on p. 7 and in note 80) is very useful in that it provides a large sample of the evidence on which one must rely in finding a MIA dialect contemporary with the Ṛgveda.

Quite sound pieces of evidence would seem to be the following. *Múhu*, *múhur* 'suddenly', *muhūrtā-* 'a moment' are connected with Avestan *mərəzu-* 'short' (in compounds), Greek *βραχύς*, Latin *brevis*; here **m]r > u* would be MIA. The connection of the *m-* and the *b-* forms has been ac-

an inheritance from pre-Sanskrit. It should be noted, moreover, that even as a Middle-Indicism a sporadic *t < t* is difficult.

¹¹ Edgerton, *Buddhist Hybrid Sanskrit Grammar* 1-2 (New Haven, 1953). For Gāndhārī Prakrit, see now J. Brough, *The Gāndhārī Dharmapada* (London, 1962).

¹² *Histoire de la langue sanskrite* 30 n. 1 (1956).

cepted in general with less than enthusiasm (*muhur*, etc., not given in WP, WH, Boisacq; given in Burrow; Mayrhofer seems to have no qualms), but even so, to connect Ṛgvedic *múhur* with the Avestan form requires the assumption of MIA phonology in the Ṛgveda. Words like *vikaṭa*- 'enormous' (< **vikṛta*-) and *śithirā*- 'loose' (< *śrth*...) have MIA phonology. So also does *pūruṣa*- 'man', which in the Ṛgveda occurs written both *pūruṣa*- and *pūruṣa*-, and in MIA shows such forms as Pali *poṣa*-, *porisa*-, BHS *poṣa*-, Pkt. *purisa*-; the old theory¹³ that all these are derived by MIA developments from **pūrṣa*- still seems most useful, even if Edgerton has qualms.¹⁴

The very frequent noun *jyōtis*- 'light' has long been recognized as having in its initial *jy*- the MIA reflex of *dy*-, viz. *j*-, with *y* written perhaps because of some thought of the various forms of the origin morpheme *dyut*- 'to be bright'; the noun in Prakrit is *joṭi*-. That the Ṛgvedic pronunciation of the noun, whether as a simplex or as first member of a compound, was *jōtis*-, is clear from a metrical examination of all the many Ṛgveda passages that could be found for it by use of Grassmann's dictionary. The stem is always disyllabic, never showing *jīy*- according to Sievers-Edgerton's law, in spite of the very great majority of its occurrences in a heavy situation (heavy 102 : light 40).¹⁵

Finally, we should note the nom.-acc. sg. neuter of the interrogative pronoun, viz. *kīm*, which P. Tedesco¹⁶ has demonstrated very beautifully to be a Middle-Indicism in the Ṛgveda. The Ṛgveda had the series *tāt*, *yāt*, *kāt*. Later, MIA had the series *taṃ*, *yaṃ*, *kiṃ*, the latter replacing **kit*, which was an analogical replacement for still earlier *cit*, which in the Ṛgveda had already become 'adverbial' and been replaced by *kāt* on the analogy of

¹³ Wackernagel 1, §51; R. Pischel, *Grammatik der Prakrit-Sprachen*, §124 (Strasbourg, 1900); Wilhelm Geiger, *Pāli Literatur und Sprache*, §30 (Strasbourg, 1916); etc.

¹⁴ *Buddhist Hybrid Sanskrit Dictionary*, s.v. *poṣa*. Ṛgvedic meter in two passages may be mended by assuming **pūrṣa*-. RV 10.90.4d *tripād ūrdhvā ūd ait pūruṣaḥ*, would be a good anuṣṭubh line with *pūrṣaḥ*. Oldenberg rather favored mending the bad meter of 10.87.16c by reading *pāuruṣeyena* as *pāurṣeyena* (*yāḥ pāuruṣeyena kraviṣā samañktē*), with the secondary derivative **pāurṣeya*- from **pūrṣa*-. He inconsistently thought it probably better to mend 10.90.4d by secondary crasis of *ūrdhvā ūd*.

¹⁵ The word is not treated in Edgerton's two papers. In addition, it is possible that the many exceptions, or some of them, to Sievers-Edgerton's law for the 'sky, day' word, when it is written with *dy*- in a heavy situation, should be explained by a MIA pronunciation *j*-. This morpheme certainly needs further study, if only to sort out the occurrences of exceptions, as Edgerton said in *Lg.* 10.253. The further note in *Lg.* 11.120 does not exhaust the matter.

¹⁶ *Lg.* 21.128-141 (1945).

kās, kā, and so on, tát, yát. Tedesco makes it clear that there must at an early period have been, beside the Ṛgvedic hieratic dialect, a colloquial with the series *tam, yam, kim*. He attempts several complicated, inconsistent, and not entirely convincing arguments as to why the whole Ṛgvedic series was not replaced by the whole MIA series. The preferable one of his arguments is perhaps that in which *tam* and *yam* were avoided as being "low" and (I suppose) "mistaken" forms for *tát* and *yát*, while *kim* was too unlike *kát* to need to be so avoided. Another possible argument is that *tam* and *yam* were avoided because of their inconvenient homonymy with the acc. sg. masculine and feminine forms (it is an inconvenient homonymy in MIA—*taṃ* and *yaṃ* for all three genders), while *kim* did not suffer from this disability and could be admitted. I prefer this type of argument, feeling that the avoidance of "low" forms might well have excluded *kim* as well as *tam* and *yam*. Neither Tedesco nor I can explain why *kim* displaced only some of the occurrences of *kát* and not all—in fact, why there was dialect mixture, even to occurrence of both forms in the same verse in three instances. Tedesco tackles the problem of nomenclature boldly and prefers to call this dialect parallel to Ṛgvedic "archaic Middle Indic"; probably "Proto-Middle-Indo-Aryan (or Indic)" is as good. No absolute chronology is possible for it, since the Ṛgvedic corpus is so impossibly vague in its chronology, beginning from the time when composition was going on in something close to the colloquial and reaching down to the time when epigonoi were composing in an archaic dialect far different from their colloquial or colloquials, some Sanskrit-like, some Proto-Middle Indo-Aryan. But we can guess that the latter type of dialect was in existence almost from the beginning of the time span.

Once this Proto-Middle-Indo-Aryan is admitted, it is not necessary to discuss further the appropriateness (or lack of it) of referring as MIA to the change of **gh*, **gh*[front vowels] and **gʷh*[front vowels], *dh*, and *bh* to *h*. For as much of the change as is contained within the time span of the Indic documents, MIA would seem to be the proper designation. The adverb *ihá* 'here', which is both Ṛgvedic and classical, is MIA and presumably was borrowed first from a PrMIA dialect, while Pali, Gāndhārī Prakrit, BHS, Śaurasēnī, and Māgadhi have *idha* (cf. Avestan *ida*), and *iha* is the form of most other Prakrit dialects, including Māhārāṣṭrī.

The net of dialects we have posited for OIA includes, then, at some rather indefinitely placed period, dialects that have already gone some way on the MIA road. At a later period all vernacular dialects are of the MIA type and, also, we must posit, form a net. This net is represented in the written record by a considerable number of literary languages.

And later, of course, comes the dialect net that we know as the NIA vernaculars, with literatures representing a dozen or so nodes in the net.

The general question arises whether we can recover from the MIA and NIA records material of Indo-Iranian or Indo-European origin which is not represented in the OIA documentary record. An *a priori* answer will not do, since it is conceivable that we could argue either way—either that our OIA records are representative of all and the only Indo-European speakers who penetrated Indic territory, or that they represent the speech of only some, perhaps even only a few, of such immigrants. Frankly, we are totally uninformed of the history of the migration. If the former linguistic hypothesis were the case, the picture would be like that of the Romance languages coming from Indo-European through the Latin channel only and adding nothing to our knowledge of Indo-European that we do not already get from the Latin record—this at least seems to be implicitly, or even more or less explicitly, the standard doctrine about the Romance languages, though it is at least possible that the picture is overdrawn and that there are a few scraps of evidence for Indo-European to be extracted from the Romance languages (so Malkiel, in conversation). The other case would be more like that of Germanic, where the literatures of the medieval period do not remotely exhaust all the languages and dialects of that period and where even scraps of other medieval evidence and much of the modern material is employable for Indo-European research.

There has been both implicit adoption of one point of view or the other, and some explicit discussion of the problem. In fact, neither point of view can be adopted in an extreme way. A conservative approach will not rule out the possibility, or even the probability, of the OIA records representing only a part of the OIA net—there is some evidence from the Dardic and Kafir languages looking in this direction. Nor will such a view fail to exercise the utmost caution in examining MIA and, *a fortiori*, NIA evidence which it is claimed represents IE material that bypassed OIA. It is *a priori* unlikely that MIA, and *a fortiori* NIA, will preserve, for example, ablaut grades that are divergent from those found in OIA, considering that the OIA ablaut structure no longer exists as a structure anywhere in MIA, and also considering how much the analogical process has remade those parts of the total OIA structure that showed most of the old ablaut system. Renou's somewhat differently slanted statement is worth quoting:¹⁷ "The time is past when one could hope to use Indo-European to explain many developments

¹⁷ *Lg.* 29.186 (1953).

peculiar to the historical period; today innovations are considered more important than survivals."

In practice, there have been divergent approaches to the problem. Research in this area has usually been of a piecemeal sort, overeager to find an Indo-Europeanism for any small problem that was being investigated. Only so can we evaluate the long lists of items collected, such as those by T. Burrow, Jules Bloch, or Wackernagel,¹⁸ these lists overlapping very considerably; the same evaluation is needed also for L. H. Gray's determined onslaught against MIA in "Fifteen Prākṛit-Indo-European Etymologies."¹⁹ Negative reactions to many of the items have been shown by many scholars, of whom I may perhaps mention Edgerton and myself. Mayrhofer alone, I think, has perforce, in the work on his etymological dictionary, had to be seriously explicit about the possibilities. His article, "Das Problem der indogermanischen Altertümlichkeiten im Mittelindischen,"²⁰ has upheld an even more strongly negative attitude than I am willing to voice.²¹

As examples where we can be sure of most of the relevant facts we can adduce *gūtha*- 'excrement' and *pard*- 'to break wind' (*pardate*). Both are found in the Sanskrit lexica, but not in the literature, except that a late commentator, who is notorious for his use of lexical words, uses the first, and a writer of poetics the second (presumably to label it a word to be avoided in literature). Both words must have been items of the vernacular vocabulary, adjudged too "vulgar" to be used in literature. Both have very good Indo-European etymologies. The first, *gūtha*-, is known from MIA and NIA, the latter, *pard*-, only from NIA. If it had not been that the Hindu grammarians and scholars constructed lexica, dredging up words from many sources besides the literature, these words would not have been known as existing in OIA dialects. What other items does MIA show that are either not attested in OIA or that are more archaic than the corresponding items in OIA? The adverb *idha*, that was adduced above, is clearly one and cannot be explained otherwise.

But hardly any other item in the lists mentioned above can be, or has been, wholeheartedly accepted. To find other than the OIA ablaut grades as explanations for MIA *supina*- 'sleep' (: Greek *ὑπνος*; contrast Skt. *svapna*-), Pali *garu*-, Pkt. *garua*- 'heavy' (: Greek *βαρὺς*; contrast

¹⁸ Burrow, *The Sanskrit Language* 45 ff.; Jules Bloch, *L'indo-aryen du Veda aux temps modernes* 14 f. (Paris, 1934); Wackernagel 1, reproduced by Renou in Introduction 8 f.

¹⁹ *JAOS* 60.361-369 (1940).

²⁰ *Studia Indologica, Festschrift für Willibald Kirfel* 219-241 (Bonn, 1955).

²¹ *Lg.* 34.409-15 (1958), 39.103-104 (1963).

Skt. *guru-*), Pkt. °*metta-* 'measure; only' (: Greek μέτρον < IE **metróm*; contrast Skt. *mātrā-*), hardly seems justified when very easy analogies are at hand in Skt. *supta-* 'asleep', Skt. *garīyāms-* 'heavier', *garīṣṭha-* 'heaviest', Pkt. *garīṭṭha-* id., MIA *mia-* (< *mita-*), *mi-* (< *miti-*), *mijjanta-*, *mijjamāṇa-* (< passive *mīya-*).²² MIA *tārisa-* 'like that' is certainly derived from OIA *tādrśa-* (whatever the explanation of *r*), and not to be compared with Greek τηλικος. Tempting as it might be, with R. L. Turner,²³ to derive MIA *ghara-* 'house' (and the derived NIA forms) from IE **gʷhoro-* 'fire, hearth (> home)', cognate with Greek θέπος, Skt. *gharma-* 'heat', it is hardly likely that *ghara-* is anything more than a metathesized form from OIA *grha-* 'house', especially since there is no trace elsewhere in this 'heat' group of etyma of the required meaning development (whatever there may be in the **aydh-* group). Even more tempting is the connection of BHS, Pkt. *se*, the third person enclitic pronoun, with the Iranian forms, Old Persian *šaiy*, Avestan *hōi*, *hē*, *šē*, and even possibly the Old Latin anaphoric pronouns *sum*, *sam*, *sos*, *sās*; and yet the extension of *s*-forms in the MIA demonstrative pronoun (from *sas*, *sā*),²⁴ and the absence of *se* in Pali make it highly probable that MIA *se* is merely an analogical formation on the model of first person *me* and second person *te*.

Pischel and Geiger²⁵ had proposed that the reason Skt. *kṣ* was variously represented in MIA by *cch*, *kkh*, *jḥ*, *ggh* was that the latter were reflexes, at least in part, of the different Indo-European combinations that fell together in Skt. *kṣ* but that were kept apart in Iranian. Geiger found that there were too many exceptions for him to accept Pischel's doctrine completely. Mayrhofer and Katre²⁶ pointed out even more exceptions, and were unwilling to accept even the points that Geiger kept, to such an extent that Mayrhofer wished to find only secondary dialectal variations and developments in the whole body of MIA instances.²⁷ H. W. Bailey accepts the older doctrine,²⁸ though he must acknowledge the exceptional character of some instances. Most recently Burrow²⁹ has combatted Mayrhofer's skepticism, especially by ingenious

²² Cf. Edgerton, *JAOS* 73.117 (on *supina-*); Emeneau, *Lg.* 39.103 (on °*metta-*); also in Mayrhofer. For *supina-* it should be noted that MIA has no forms in the verb with any vowel but *u/o*.

²³ *BSOS* 3.401-404 (1924). But Turner now (*A Comparative Dictionary of the Indo-Aryan Languages* 239, entry 4428) says that his earlier view is "very unlikely," apparently persuaded by Mayrhofer's skepticism (1.357).

²⁴ Pischel, §423; Edgerton, *BHS Grammar*, chap. 21, esp. gen. pl. *sānam* in 21.45.

²⁵ Pischel, §§317-326; Geiger, §56.

²⁶ Mayrhofer, *Studia Indologica* . . . *Kirfel* 227-233; Katre, *JBORS* 23.15 (1937).

²⁷ Edgerton, *BHS Dictionary*, s.v. *pragharati*, shows skepticism of the older doctrine.

²⁸ *BSOAS* 13.137 (1949).

²⁹ *JAOS* 79.255-262 (1959), esp. 261 f.

new etymological combinations and philological interpretations. He probably has made advances, but even so he is forced to posit *ad hoc* interdialectal borrowings and influences on MIA from Sanskrit, and the end result is probably too complicated to be really cogent. In my opinion, the matter is still *sub judice*; my guess is that Mayrhofer's skepticism is better justified than the old doctrine.

If MIA adds only a very little that is certain to the OIA picture of Indo-European inheritances in Indic, it can hardly be expected that NIA will contribute much, and in fact there are few suggestions. Hindi *ālā* 'meal', and so forth, and the fourteenth-century A.D. Sanskrit lexical *ālta* 'food', which is surely only a Sanskritization of an NIA form, are hardly to be derived from Indo-Iranian **ar-* (IE **al-*) 'to grind', which otherwise has no representative in Indo-Aryan.³⁰

The most notable of such suggestions is that the past passive stem of the verb meaning 'to give', as seen in Hindi *diyā* and other NIA forms, is derived from IE **dātō-*, through **dita-* of an unattested OIA dialect and such an unattested MIA form (with an added OIA *-ka-*) as **diaa-*; so Tedesco³¹ argues most ingeniously and brilliantly and with much reconstruction to which one can take no exception. His central thesis, however, has been much discussed, and independently (so I think) in recent years has been rejected by three scholars—Mayrhofer, myself, and Turner.³² The two latter find in the NIA forms analogical formations based on a present stem of the type *dē-*, which is found early in MIA

³⁰ *Kurzgefasstes Etymologisches Wörterbuch des Altindischen* 1.546 (Heidelberg, 1956).

³¹ "Geben und Nehmen im Indischen," *JAOS* 43.358-390 (1923). In saying that **dita-* is unattested in OIA, I am not ignoring Epic Skt. *vyādita-* 'opened' (*vi-ā-√dā*), which has been discussed often. It is generally agreed (e.g., by M. Leumann *IF* 57.231, and Wackernagel-Debrunner, *Altindische Grammatik* 2.2.561 f. [Göttingen, 1954]) that this form can be only a late formation of MIA type, on the basis of a present **vy-ā-deti*. What seems not to have been put in evidence so far is that the form is found only in the compound *vyādītāsya-* 'with opened mouth' and usually only in the simile *vyādītāsya ivā 'natakāḥ* (Mbh. 6.58.33d), *vyādītāsyam ivā 'ntakam* (6.59.20d, 102.18b, 110.38d, 112.74f; according to E. W. Hopkins, *The Great Epic of India* 436 [New York, 1901], also in *Rāmāyaṇa*, Bombay ed., 3.2.6); otherwise *vyādītāsyair mahānādaiḥ* (Mbh. 2.22.23a) and *vyādītāsyā mahāraudrā* (3.146.46c). The synonymous equivalent of the simile, *vyāntānanam ivā 'natakam*, is found in Mbh. 3.125.1d, 6.55.45b, 59.23b, 78.23d, 103.93b, 104.37f (and also in *Rāmāyaṇa* according to Hopkins). Nothing at the moment seems to follow from this co-existence of synonymous compounds, one with the Pāṇinian form and one with the other problematic one; nor, of course, are the collectanea complete for the epic at the present stage of epic studies (probably only for Mbh. books 1, 2, 3, 6).

³² Mayrhofer, *Studia Indologica* . . . *Kirfel* 237-239, and *Kurzgef. etym. Wb.* 2.13 f. (my interpretation is that he has rejected it); Emeneau, *Lg.* 34.409 f. (1958); Turner, *Journal of the Gypsy Lore Society*, 3rd ser., 39.28-30 (1960), in spite of earlier acceptance, e.g., in *BSOS* 5.131 (1928).

and which (as I stated it) led to the analogy *eti* 'goes' : *ita-*, *jeti* 'conquers' : *jita-*, (s)*theti* 'stands' : *sthita-*, *dheti* 'places' : *hita-*, *deti* 'gives' : *x*, *x* being **dita-*.

Not quite all the data have yet been included by any of those who have treated the matter. Evidence for the NIA languages of the northern section of the Indo-Iranian border, the Dardic and Kafir languages, is still somewhat scanty, and not all the material relevant to this problem is yet in hand. However, Khowar (otherwise Chitrali), one of the Dardic languages,³³ provides some interesting forms. For the verb 'to give', the perfect participle is *dirū*, which Morgenstierne derives from **ditaka-*, which is the OIA origin of MIA **diaa-*, whence Hindi *diyā* (**t* > Khowar *r* intervocalically). It is hardly necessary to separate this participle in origin from those of other NIA languages already presented. It is interesting, however, that the *r* of Khowar is independent evidence for **t*, which those NIA languages that have descended through the normal MIA languages do not actually show (Hindi *diyā*, etc.). Since we have posited that this participle is analogically based on a present stem *dē-*, we should examine the Khowar verb to try to determine whether our explanation will hold for it, too. The Khowar present (really present and future, labeled 'aorist-future') is *dōm*, *dōs*, *dōi*, *dōsi*, *dōmi*, *dōni*. The Dardic languages in general, however, have a stem *dē-*.³⁴ The stem *dō-* of Khowar must be analogically formed on the basis of the small class of verbs that have this type of stem, cf. *korōm* 'I do' (< Skt. *karōmi*), *bōm* 'I become' (Skt. *bhavāmi* > MIA [e.g., Pali] **bhōmi* > Khowar *bōm*); it is not of an origin parallel to that of the Balto-Slavic **dōmi*, *dōsi*, and so on (cf. Senn's paper, p. 149 of this volume). Etymologies are uncertain or unknown for the other verbs of the class. We may posit, then, for Khowar the same formations that enter into the analogy given above, and should probably assume that, at the early MIA period when the analogical formation took place, the ancestor of Khowar formed part of the MIA dialect net.

The general run of NIA languages thus seems unpromising in yielding material of the sort we are looking for. It has been claimed that the Dardic and Kafir languages already mentioned occupy a different position in NIA, and it is possible that when we have more descriptive data on them there may be found, rather abundantly, Indo-European evidences that appear in neither the OIA nor the MIA records. It was claimed by Grierson (as well as by some before him) that these two

³³ G. Morgenstierne, "Some Features of Khowar Morphology," *Norsk Tidsskrift for Sprogvidenskap (NTS)* 14.5-28 (1947). Some lexical material is presented also by him in an article, "Sanskritic Words in Khowar," *Felicitation Volume presented to Prof. Sripad Krishna Belvalkar* 84-98 (Banaras, 1957).

³⁴ Turner, *Dictionary of the Nepali Language* (London, 1931), s.v. *dinu*.

groups of languages form a third branch of Indo-Iranian, in that "they seem to have left the parent stem after the Indo-Aryan languages, but before all the typical Iranian characteristics, which we meet in the Avesta, had become developed."³⁵ The material that was gathered by Morgenstierne after Grierson's volume appeared led Morgenstierne to the conclusion (which has been accepted by, for example, J. Bloch and Burrow³⁶) that the Dardic languages (Kashmiri, Shina, Indus Kohistani, Khowar, Kalasha, Pashai, Tirahi) are Indo-Aryan but did not pass through the MIA developments represented by the records, while, on the other hand, the Kafir languages (Kati, Waigali, Ashkun, Prasun, and to some extent Dameli) may occupy some sort of special position.³⁷ The task of sorting out the evidence is considerably complicated by loanwords in the Kafir languages from neighboring Iranian languages and from other neighboring Indo-Aryan languages, and also by loans in the other directions, that is, from the Kafir languages into neighboring Iranian and Indo-Aryan languages. But, neglecting such items (insofar as they can be identified), one finds that in the Kafir languages there is retention with Indo-Aryan of *s*, which Iranian changes to *h*, and retention with Iranian of two separate palatal series, which Indo-Aryan let fall together. I.e., **ǵ*, *ǵh* > Kaf. *j* (= *dz*)/*z*; palatalized **g* and *gʷ*, *ǵh* and *gʷh* > Kaf. *j̃/ž*; e.g., Kati *j/zǎ*, Waigali, Ashkun *zǎ* 'knee'; Kati *j/zim*, Waigali *zim*, Prasun *zəma*, Ashkun *žim* (secondary palatalization before *i*) 'snow'; Kati, Waigali *žī*, Prasun, Ashkun *žī* 'bowstring'; Kati *jǎř-*, Waigali *jǎ-* 'to kill'. In addition, there is the following feature in which the Kafir languages are more archaic than either Iranian or Indo-Aryan, viz., in having to a large extent affricates (dental) for the Indo-European palatal series, even though the matter is at times obscured by borrowings from Indo-Aryan, especially in Waigali and Ashkun. Examples: Kati *duč* 'ten', Prasun *leze* id., *čpu-lč* 'fourteen', Waigali *dōš*, Ashkun *dus* 'ten'; Kati *cāw*, Waigali, Prasun *cāw*, Ashkun *c/sau* 'branch'; Kati *čūš*, Waigali *čūne-*, Ashkun *čuni-* 'empty'; and 'knee' and 'snow'

³⁵ G. A. Grierson, *The Pisāca Languages of North-western India* . . . , p. iii (London, 1906). Sköld held that the Kafir languages are not Indo-Iranian at all, but an independent branch of Indo-European (so reported by Morgenstierne, *NTS* 2.196, with reference in note 1 to an unavailable paper by Sköld). This is surely fantasy.

³⁶ Bloch, *L'indo-aryen* 18 f.; Burrow, *The Sanskrit Language* 32.

³⁷ Morgenstierne, *Report on a Linguistic Mission to Afghanistan* (Oslo, 1926), *Report on a Linguistic Mission to North-western India* (Oslo, 1932), "The Language of the Ashkun Kafirs," *NTS* 2.192-289 (1929), "Additional Notes on Ashkun," *NTS* 7.56-115 (1934), "Indo-European *k* in Kafir," *NTS* 13.225-238 (1945), "The Language of the Prasun Kafirs," *NTS* 15.188-334 (1949), "The Waigali Language," *NTS* 17.146-324 (1954). Recent fieldwork by Georg Buddruss in this area may have included the Kafir languages, as it did Dardic, but it has not yet been published.

above.³⁸ Morgenstierne argues for another archaism in the occurrence of *s* after original *u* (e.g., Kati *dūs*, Waigali, Ashkun *dōs*, Prasun *ulus* 'yes-terday': Dardic *dōṣ* id.: Skt. *doṣā* 'in the evening'; Kati *mūsə*, Ashkun *musā*, Prasun *mūsu* 'mouse'), and *š* after *i* (with secondary palatalization; e.g., Kati *wiš* 'poison', *niš-* 'to sit down'), as against *ś* initial (Kati, Waigali, Ashkun *šū*, Prasun *wuṣu* 'six') and after *r* (Kati, Waigali *waš* 'rain'). But this distribution might equally well result from conditioned development at a late period (unless there are reasons, unknown to me, for thinking otherwise); there are exceptions to the conditions just stated, but either interpretation will divide the material in the same way.

In general, the Kafir vocabulary agrees with Indo-Aryan rather than with Iranian. There are a very few instances of agreement with Iranian; e.g., Kati, Waigali *kan-*, Ashkun *kōn-* 'to laugh' (Persian *xand-*).

Morgenstierne's final view on the position of Kafiri³⁹ is that "the remarkable archaisms of Kaf. and its geographical position render it probable that it contains a residuum going back to the language of tribes which split off from the main body of Aryans and penetrated into the Indian borderland before the invasion of the Indo-Aryans." Kafiri is in the main Indo-Aryan in vocabulary and phonology, but it retains the pre-Indo-Aryan treatment of the two palatal series, and, if Morgenstierne is correct in his view of the relationship of the Kafiri dental affricates to the Indo-European palatal series, it represents a very archaic state of things in this matter, perhaps a stage of development that may be called Proto-Indo-Iranian.

Close examination of the Dardic and Kafir languages has already yielded to Morgenstierne more details than I have reported here, of a very archaic layer of vocabulary. For example, he gives Khovar *bispī* or *bispiki*, Waigali *vašpik* 'wasp', which is represented in Iranian by Baloch *gwabz*, but is not otherwise found in Indo-Iranian. Some of the other suggestions are not equally convincing. Further collection of field material to supplement the fairly meager collections already made, and further detailed study by Iranianists and Indo-Aryanists, will undoubtedly yield more material of the type we have been interested in here. The Indo-Iranian border area is both neglected and of extreme interest. Without its evidence the dialect jigsaw-puzzle picture, whether of Indo-Aryan or of Iranian, shows many missing pieces.

(Participants in the discussion following the conference presentation of the first version of this paper: Puhvel, Hoenigswald, Lane, Lehmann, Polomé, Hamp.)

³⁸ Morgenstierne argues that this is an archaism against M. Leumann (*IF* 58.3 [1941]), who would find in this material a change from sibilant to affricate in some dialects. Morgenstierne makes the good point (*NTS* 13.230) that no instances of IE **s* show *č* in Kafiri, except in a few cases of assimilation.

³⁹ *NTS* 13.234 (1945).